Whole-genome Duplications and the Long-term Evolution of Gene Regulatory Networks in Angiosperms

Fabricio Almeida-Silva [1,2] and Yves Van de Peer [*,1,2,3,4]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

[2]VIB Center for Plant Systems Biology, VIB, Ghent, Belgium

[3]Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

[4]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, China

***Corresponding author:** E-mail: yves.vandepeer@psb.vib-ugent.be.

Associate editor: Dr. Michael Purugganan

Abstract

Angiosperms have a complex history of whole-genome duplications (WGDs), with varying numbers and ages of WGD events across clades. These WGDs have greatly affected the composition of plant genomes due to the biased retention of genes belonging to certain functional categories following their duplication. In particular, regulatory genes and genes encoding proteins that act in multiprotein complexes have been retained in excess following WGD. Here, we inferred protein–protein interaction (PPI) networks and gene regulatory networks (GRNs) for seven well-characterized angiosperm species and explored the impact of both WGD and small-scale duplications (SSDs) in network topology by analyzing changes in frequency of network motifs. We found that PPI networks are enriched in WGD-derived genes associated with dosage-sensitive intricate systems, and strong selection pressures constrain the divergence of WGD-derived genes at the sequence and PPI levels. WGD-derived genes in network motifs are mostly associated with dosage-sensitive processes, such as regulation of transcription and cell cycle, translation, photosynthesis, and carbon metabolism, whereas SSD-derived genes in motifs are associated with response to biotic and abiotic stress. Recent polyploids have higher motif frequencies than ancient polyploids, whereas WGD-derived network motifs tend to be disrupted on the longer term. Our findings demonstrate that both WGD and SSD have contributed to the evolution of angiosperm GRNs, but in different ways, with WGD events likely having a more significant impact on the short-term evolution of polyploids.

Key words: **systems biology, genome evolution, polyploidy, network evolution, bioinformatics.**

Introduction

Whole-genome duplications (WGDs) are a key source of extra genetic material for evolution to work with (Ohno 1970), and they have occurred independently in multiple branches of the tree of life. For instance, vertebrates have experienced two rounds of WGD early in their evolution (Dehal and Boore 2005; Van de Peer et al. 2010; Nakatani et al. 2021), with an additional WGD event shared by all teleosts (Amores et al. 1998; Taylor et al. 2003). The genomes of several yeasts have also undergone ancient WGD events (Wolfe and Shields 1997). However, WGD events have been most abundant in plants, and they are thought to have contributed to the radiation of important plant families (Schranz et al. 2012; Tank et al. 2015; Landis et al. 2018), to increased diversity (Ren et al. 2018), and to survival in stressful times (Van de Peer et al. 2017; Van de Peer et al. 2021).

After a WGD event, survival and establishment of newly formed polyploids is challenging due to the detrimental effects of doubling the entire set of chromosomes, which leads to for example genomic shock and reduced fertility (Comai 2005; Woodhouse et al. 2010). Polyploids that survive a WGD event typically undergo a rediploidization process that leads to genome fractionation (i.e., loss of functional DNA sequences), probably because of functional redundancy (De Smet et al. 2013; Li et al. 2016). During the process of rediploidization and fractionation, it has been observed that genes encoding proteins involved in intricately connected systems, such as transcription factors, kinases, and members of multiprotein complexes, tend to be retained more often than other genes after WGD (Blanc and Wolfe 2004; Maere et al. 2005; Freeling 2009). To date, this biased retention of particular gene classes is best explained by the gene balance hypothesis, which states that some gene families are dosage sensitive; hence, their retention preserves stoichiometric balance (Freeling and Thomas 2006; Birchler and Veitia 2010). For the same reason, retention of such dosage-sensitive genes is usually selected against when these are duplicated in small-scale duplication (SSD) events, as stoichiometric balance is disrupted (Maere et al. 2005; Freeling 2009).

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7221... on 29 August 2023

Article

The preferential retention of transcription factors (TFs) following WGDs motivated researchers to explore the impact of WGD events on the evolution of the TF repertoires and gene regulatory networks (GRNs) (De Smet and Van de Peer 2012; Panchy et al. 2019; Moharana and Venancio 2020; Mottes et al. 2021; Gera et al. 2022). In particular, WGD can reorganize the topology of GRNs by creating novel network motifs, which are enriched subgraphs typically found in complex networks (Milo et al. 2002). Network motifs are simple genetic circuits regarded as the building blocks of biological networks, and the connections that occur within and between motifs result in robust regulatory interactions observed in complex biological systems (Milo et al. 2002; Burda et al. 2011). As motifs perform elementary regulatory functions, they have been positively selected during the evolution of biological networks (Ward and Thornton 2007; Mottes et al. 2021), and examples include V motifs (paralogous TFs regulating a shared target gene), lambda motifs (paralogous target genes regulated by a shared TF), and bifans (a pair of paralogous TFs that regulate the same pair of targets) (Mottes et al. 2021).

Analyzing motif frequencies can reveal the impact of gene and genome duplications in biological networks and thus evolutionary innovation or (increase in) biological complexity. For instance, in human GRNs, WGD events have increased regulatory redundancy (and, hence, increased mutational robustness) and generated complex combinations of motifs, which enhanced the performance of high-level functions such as signal integration and noise control (Mottes et al. 2021). Likewise, WGD events led to the enrichment of bifans in yeast, which have been shown to be important for signal processing and formation of modular network structures (Ward and Thornton 2007). However, little is known about the impact of WGD on the topology of GRNs in angiosperms. As angiosperms have a more diverse WGD background, both with respect to the number and age(s) of WGD events across clades, studying the evolution of GRN topologies will further our understanding on the significance of WGDs—versus SSDs—for the evolution of flowering plants.

Here, we explored the impact of WGD and SSD events on protein–protein interaction (PPI) networks and GRNs of seven angiosperm species with well-defined WGDs. We observe that PPI networks are enriched in WGD-derived genes associated with intricately connected molecular processes. We show that interacting WGD-derived gene pairs have lower rates of sequence divergence and higher preservation of interactions as compared to SSD-derived gene pairs. Besides, genes originating from recent WGD events are more often part of GRN motifs than genes from ancient WGD events, and recent polyploids generally have higher motif frequencies. WGD events are the main source of novel motifs for recent polyploids, suggesting an immediate selective advantage for polyploids under certain conditions. However, WGD-derived motifs are lost over time and after adaptation. Finally, WGD-derived genes in GRN motifs seem associated with dosage-sensitive processes, such as transcriptional regulation, cell cycle, histone modifications, flower development, and photosynthesis, whereas SSD-derived genes in motifs seem associated with response to biotic and abiotic stress.

Materials and Methods

Genomic Data Acquisition

We obtained genomic data for seven angiosperm species: *Glycine max*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Populus trichocarpa*, *Vitis vinifera*, *Zea mays*, and *Oryza sativa* (table 1). Protein sequences, DNA sequences for coding regions, and genome annotations were downloaded from Ensembl Plants release 53 (Yates et al. 2022) (supplementary table S1, Supplementary Material online). RNA-seq data from baseline experiments were downloaded from European Bioinformatics Institute's Expression Atlas (Papatheodorou et al. 2020) (supplementary table S1, Supplementary Material online). Whole-genome functional annotations (Gene Ontology terms and InterPro domains) were obtained from Ensembl's BioMart using the R package biomaRt (Durinck et al. 2009).

Prediction and Classification of TFs

The R package *planttfhunter* (https://bioconductor.org/packages/planttfhunter) was used to predict TFs from protein sequences and classify them into families using the same classification scheme implemented in PlantTFDB (Tian et al. 2020). Briefly, profile hidden Markov models for DNA-binding domains and auxiliary domains were searched against the entire proteome using HMMER 3.0 (Finn et al. 2011), and TF families were identified based on the presence and number of signature domains.

Identification of Duplicated Genes and Calculation of Substitution Rates

Duplicated gene pairs were identified and classified into WGD-derived and SSD-derived gene pairs using the R/Bioconductor package *doubletrouble* (https://bioconductor.org/packages/doubletrouble). In short, all-against-all intragenomic DIAMOND protein similarity searches (sensitive mode, e-value = 1e−10, top hits = 5) were performed to identify the entire set of paralogous gene pairs for each

Table 1. Included Species and Summary Statistics.

Species	Genes	WGD	SSD	GRN (M)	PPI (M)
Glycine max	55897	34570	83219	2.27	1.38
Arabidopsis thaliana	27628	4329	43154	2.50	0.55
Solanum lycopersicum	34429	3906	57664	1.34	0.52
Populus trichocarpa	41335	15400	63613	1.28	0.64
Vitis vinifera	35134	2860	62571	1.21	0.21
Zea mays	39756	8095	69611	2.05	0.71
Oryza sativa	35775	3435	49498	1.80	0.25

Genes, number of genes in each genome. WGD and SSD, number of WGD- and SSD-derived gene pairs. GRN and PPI, number of edges (in millions) in each gene regulatory network and protein–protein interaction network, respectively.

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

species. Paralogous gene pairs in the same syntenic blocks were classified as WGD-derived pairs, and all other pairs were classified as derived from SSD. Rates of synonymous and nonsynonymous substitutions per substitution site (K_s and K_a, respectively) were calculated using the pipeline described in (Qiao et al. 2019), which consists in performing pairwise alignments with MAFFT (Katoh and Standley 2013) followed by K_a and K_s calculation with KaKs_Calculator 2.0 using the modified Yang-Nielsen model (Wang et al. 2010). Peaks in whole-paranome K_s distributions were identified with the R package *doubletrouble* by fitting Gaussian mixture models. As Gaussian mixture models can overestimate the number of peaks in K_s distributions, we explicitly defined the number of peaks for each species based on literature evidence (Qiao et al. 2019).

PPI and GRN Inference

PPI networks were obtained from the STRING database using a confidence threshold of 0.5 (Szklarczyk et al. 2021). GRNs were inferred using the GENIE3 algorithm (Huynh-Thu et al. 2010) as implemented in the R package BioNERO (Almeida-Silva and Venancio 2022). GRNs inferred with GENIE3 were represented as a fully connected bipartite graph. To remove spurious edges, we used the function *grn_filter()* from BioNERO to rank edges and split each GRN in 20 subnetworks of increasing number of edges, where the first subnetwork contains the 0.05 quantile and the 20th subnetwork is the original fully connected graph. For each subnetwork, the function *grn_filter()* fitted the degree distribution to a scale-free topology. The largest subnetwork with an $R^2 > 0.75$ (coefficient of determination for the scale-free topology fit) was selected as the optimal GRN (supplementary fig. S1, Supplementary Material online).

Network Topology Analyses and Functional Enrichment

To ensure that the topology of the inferred GRNs and PPI networks resembled the topology of real-world biological networks, all networks were checked for scale-free topology fit using the function *check_SFT()* from the R package BioNERO (Almeida-Silva and Venancio 2022). Degree distributions for WGD- and SSD-derived gene pairs were calculated with the R package igraph (Csardi and Nepusz 2006) and visualized with the R package ggstatsplot (Patil 2021). Enrichment analyses were performed with the function *enrichment_analysis()* from the R package BioNERO. *P* values from Fisher's exact tests were adjusted for multiple comparisons with Benjamini–Hochberg correction.

Identification of Network Motifs, Significance Assessment, and Interaction Similarity

To identify and analyze graph motifs, we developed an R package named *magrene*, which is available on Bioconductor (https://bioconductor.org/packages/magrene), and the source code of which is available on a GitHub repository (https://github.com/almeidasilvaf/magrene). This package can be used to identify five types of motifs containing paralogous gene pairs, namely V, PPI V, lambda, delta, and bifan motifs (fig. 1). For each species, we calculated the frequency of motifs containing WGD- and SSD-derived gene pairs. To account for the effect of "gene age" on motif frequencies, we classified duplicated gene pairs into age groups based on K_s peaks using the function *split_pairs_by_peak()* from the *doubletrouble* package. Briefly, age groups were defined based on the mean $\pm$ 2 standard deviations relative to each K_s peak, and duplicate pairs (either WGD- or SSD-derived) with K_s values within a particular age group were assigned to it. Then, WGD versus SSD comparisons were only performed for paralogs belonging to the same age group. To assess the significance of the observed motif frequencies, we used the function *generate_nulls()* from *magrene* to generate 1000 degree-preserving simulated networks through node label permutation and calculate motif frequencies in each permutation. Then, the function *calculate_Z()* was used to obtain Z-scores for the observed motif frequencies. The function *sd_similarity()* was used to calculate the interaction similarity between gene pairs, which is represented by the Sorensen–Dice similarity index:

$$S(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

where A and B are the interacting partners of nodes A and B, respectively.

Results and Discussion

Paralogous Gene Pairs, Summary Statistics, and Network Topology

We identified and classified duplicated genes in the genomes of seven angiosperm species: *G. max*, *A. thaliana*, *S. lycopersicum*, *P. trichocarpa*, *V. vinifera*, *Z. mays*, and *O. sativa* (table 1). These species were selected to account for a broad range of plant families and WGD histories. Expectedly, all species have more SSD-derived gene pairs than WGD-derived gene pairs, but such difference in frequency is smaller for species that have undergone recent WGD events (table 1). Peaks in K_s distributions were used to ascribe duplicate gene pairs to their respective age groups (table 2; supplementary fig. S2, Supplementary Material online; see Materials and Methods for details). All WGD and SSD comparisons presented hereafter were only performed with genes from the same age group, unless stated otherwise.

The degree distributions of all inferred GRNs and PPI networks satisfied the scale-free topology fit, indicating that they have the topological properties of real-world biological networks. As differences in motif frequencies can be due to differences in node degree, we compared the degree distributions of WGD- and SSD-derived genes in GRNs and PPI networks to detect potential biases. We

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

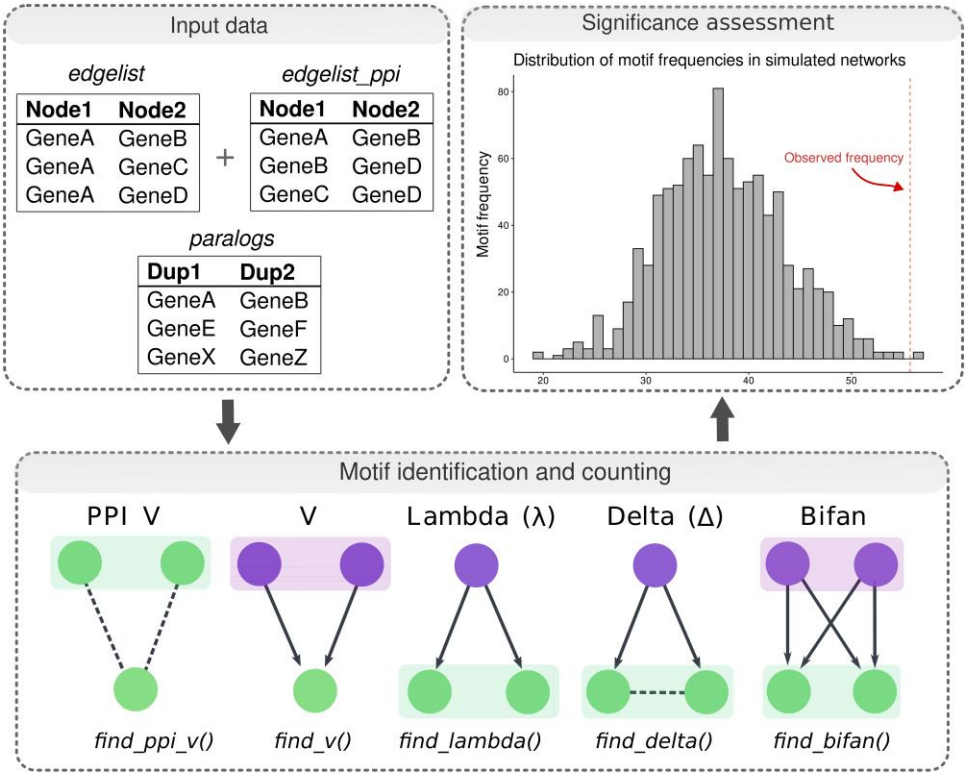

Fig. 1. Workflow of motif identification and significance assessment with *magrene*. Input data types for motif detection, which include a data frame of GRN edges, a data frame of PPI network edges, and a data frame of paralogous gene pairs. Motif types and R functions in *magrene* to find specific motifs are indicated in the bottom panel. Shaded boxes indicate paralogous gene pairs. Regulators and targets are indicated in purple and green, respectively. Arrows indicate directed regulatory interactions, whereas dashed lines indicate protein–protein interactions. Observed motif frequencies for each motif type are then compared with a null distribution obtained from simulated networks, as indicated in the histogram. See Materials and Methods for details.

Table 2. Summary of Identified K_s Peaks and Their Corresponding Whole-genome Duplication/Triplication Events.

Species	K_s peaks	Event	Event age (Ma)
Glycine max	2	Glycine WGD	13.59 (11.87–13.99)
		Fabaceae WGD	61.52 (56.04–67)
Arabidopsis thaliana	2	α WGD	50.07 (49.27–50.99)
		γ WGT	125 (117–133)
Solanum lycopersicum	2	Solanaceae WGD	61.16 (57.47–64.84)
		γ WGT	125 (117–133)
Populus trichocarpa	3	Salicaceae WGD	34.73 (32.60–36.34)
		γ WGT	125 (117–133)
		ε WGD	308 (297–319)
Vitis vinifera	1	γ WGT	125 (117–133)
Zea mays	3	Zea WGD	20.40 (19.71–20.99)
		ρ WGD	66.23 (63.08–69.89)
		σ WGD	114 (110–117)
Oryza sativa	2	ρ WGD	66.23 (63.08–69.89)
		σ WGD	114 (110–117)

Peaks were identified in whole-paranome K_s distributions. Age (in Ma) and 90% confidence intervals were obtained from (Vanneste et al. 2014; Clark and Donoghue 2017; Van de Peer et al. 2017).

observed no significant differences in degree distributions for WGD- and SSD-derived genes in both networks (Mann–Whitney *U* test; $P < 0.05$). Although some comparisons had significant differences (e.g., PPI network of *A. thaliana* and GRN of *G. max*), the small *P* values are likely due to large sample sizes, as the effect sizes are negligible (rank-biserial correlation < 0.15) (supplementary figs. S3 and S4, Supplementary Material online).

PPI Networks Are Enriched in WGD-derived Genes

We performed a Fisher's exact test using all genes as background to test for associations between duplication modes and the presence of PPIs. We observed an enrichment of WGD-derived genes in the PPI networks for all species ($P < 0.001$), suggesting that strong selection pressures constrain the protein products of WGD-derived genes to interact physically. Functional enrichment analyses revealed that the set of interacting WGD-derived genes is enriched in genes associated with signal transduction, lipid metabolism, carbohydrate and amino acid oxidation, cell wall biogenesis, redox homeostasis, translation, and transcriptional regulation (supplementary table S2, Supplementary Material online).

These findings are in line with the gene balance hypothesis, as interacting WGD-derived pairs are enriched in dosage sensitive biological processes, which tend to be retained more often after WGD events to preserve stoichiometry (Maere et al. 2005; Freeling and Thomas 2006; Freeling 2009; Birchler and Veitia 2010; Teufel et al. 2019; Almeida-Silva et al. 2020). Single-gene duplications for genes involved in complex PPI networks would likely be deleterious, as it leads to dosage imbalance. Thus, selection tends to retain such genes only if the whole PPI network is duplicated, as it happens in WGD events (Maere et al. 2005).

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

Selective Pressures Constrain Sequence Divergence in Interacting WGD-derived Genes

To understand how evolution shaped the functional divergence of WGD- and SSD-derived gene pairs, we explored how duplicates diverge in sequence over time. As synonymous substitutions accumulate in a neutral manner, K_s was used as a proxy for time, and K_a was used to represent sequence divergence. To account for saturation at higher K_s values, we fitted Michaelis–Menten curves to the scatter plot, as previously described (Tasdighian et al. 2017; Defoort et al. 2019). For all species, we found that interacting WGD-derived gene pairs have slower rates of sequence divergence over time as compared to SSD-derived genes (fig. 2).

Sequence divergence leads to novel and/or specialized gene functions, which can disrupt gene balance. Thus, the lower sequence divergence for interacting WGD-derived genes is predicted by the gene balance hypothesis. Similar findings have been observed in *A. thaliana* (Tasdighian et al. 2017; Defoort et al. 2019), *Saccharomyces cerevisiae* (Fares et al. 2013), and tomato and maize (Defoort et al. 2019). Collectively, these findings suggest that this is a universal pattern for WGD-derived genes, as it has been observed in multiple independent taxa and studies.

WGD-derived Gene Pairs Tend to Interact With the Same Partners

For each species and age group, we calculated the interaction similarity of paralogous gene pairs using Sorensen–Dice indices, a metric that ranges from 0 to 1, where 0 means no shared partners, and 1 means complete overlap of partners. Overall, WGD-derived gene pairs have a higher interaction similarity in PPI networks than SSD-derived pairs (fig. 3). The difference is even more pronounced for older duplicate pairs, with ancient WGD-derived pairs displaying a much higher interaction similarity than ancient SSD-derived pairs (fig. 3). However, we observed no difference in interaction similarity between WGD- and SSD-derived pairs in GRNs (supplementary figs. S5 and S6, Supplementary Material online).

The observed patterns in PPI networks suggest that strong selection pressures constrain WGD-derived pairs to preserve their interactions over time, whereas more relaxed pressures allow SSD-derived pairs to lose and gain interaction partners. The higher preservation of interacting partners for WGD-derived gene pairs was previously observed in *A. thaliana* and human PPI networks (Defoort et al. 2019; Mottes et al. 2021), and it can again be explained by dosage balance constraints. As SSD-derived genes are generally not dosage sensitive, they can establish novel connections in PPI networks. On the other hand, novel connections for WGD-derived genes are likely disadvantageous, because they can disrupt the stoichiometric balance in the cell.

(Recent) WGDs Fuel(ed) the Emergence of Network Motifs

We identified and counted network motifs for each species and age group using the R package *magrene* (see Materials

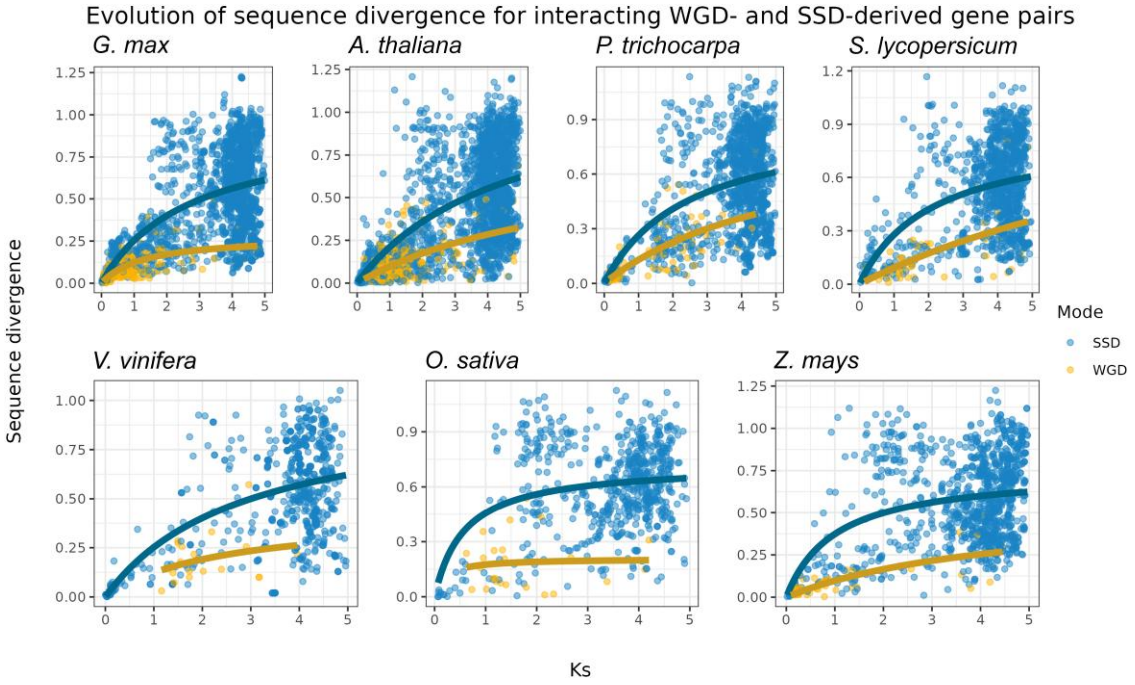

Fig. 2. Evolution of sequence divergence for interacting WGD- and SSD-derived gene pairs. Rates of nonsynonymous substitutions per substitution site (K_a) were used to represent sequence divergence, and rates of synonymous substitutions per substitution site (K_s) were used as a proxy for time. As linear models cannot account for saturation at higher K_s values, a Michaelis–Menten curve was fitted in the scatter plot. The plot shows that WGD-derived gene pairs have lower sequence divergence over time as compared to SSD-derived pairs.

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

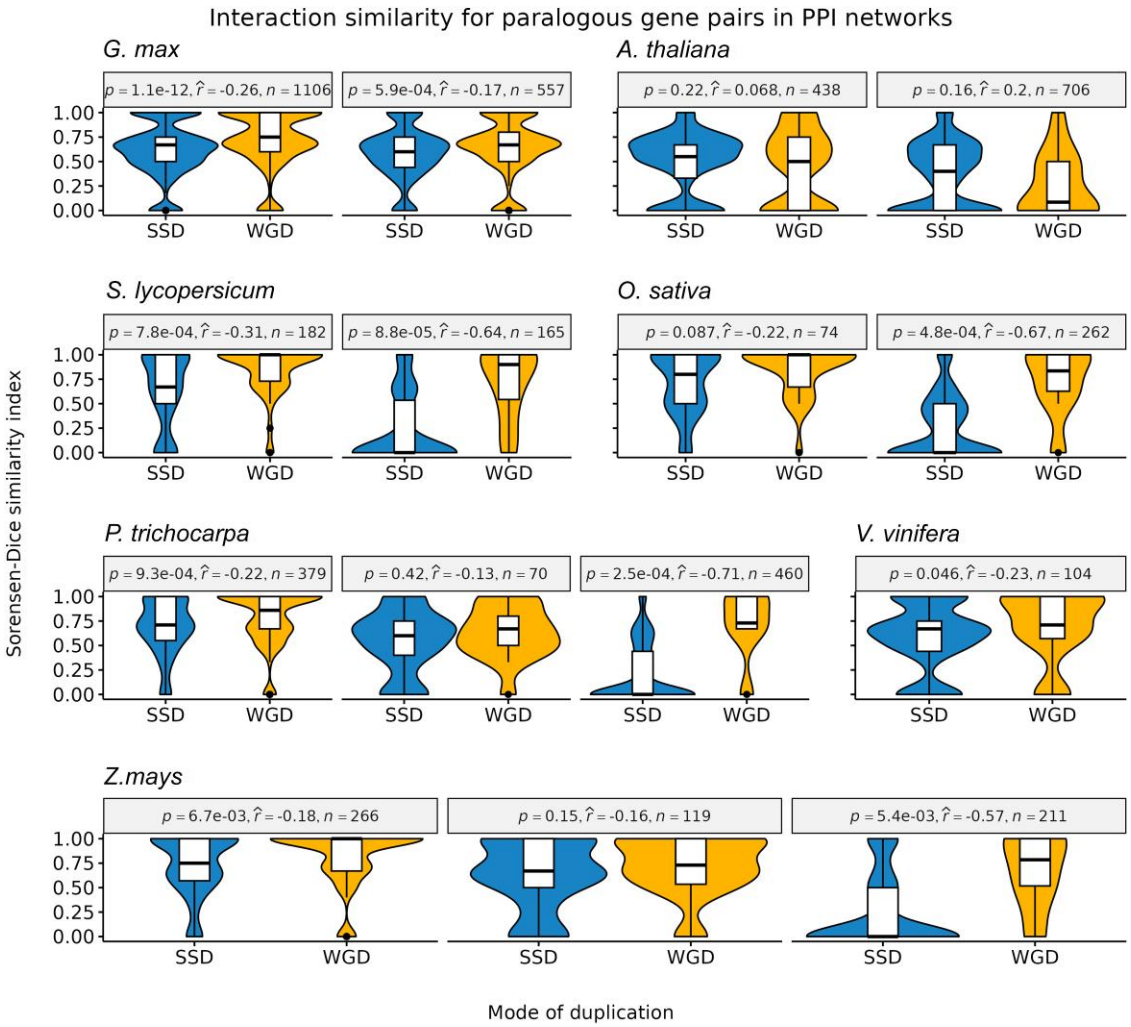
Interaction similarity for paralogous gene pairs in PPI networks
G. max
p = 1.1e-12, r̂ = -0.26, n = 1106
p = 5.9e-04, r̂ = -0.17, n = 557
A. thaliana
p = 0.22, r̂ = 0.068, n = 438
p = 0.16, r̂ = 0.2, n = 706
S. lycopersicum
p = 7.8e-04, r̂ = -0.31, n = 182
p = 8.8e-05, r̂ = -0.64, n = 165
O. sativa
p = 0.087, r̂ = -0.22, n = 74
p = 4.8e-04, r̂ = -0.67, n = 262
P. trichocarpa
p = 9.3e-04, r̂ = -0.22, n = 379
p = 0.42, r̂ = -0.13, n = 70
p = 2.5e-04, r̂ = -0.71, n = 460
V. vinifera
p = 0.046, r̂ = -0.23, n = 104
Z.mays
p = 6.7e-03, r̂ = -0.18, n = 266
p = 0.15, r̂ = -0.16, n = 119
p = 5.4e-03, r̂ = -0.57, n = 211
Sorensen-Dice similarity index
1.00
0.75
0.50
0.25
0.00
SSD
WGD
Mode of duplication

FIG. 3. –Sorensen–Dice similarity indices were used to indicate interaction similarity. For each species, different panels indicate gene pairs from different age groups (defined based on K_s peaks, see Materials and Methods for details), with panels arranged from left to right in chronological order (from younger to older groups). Overall, WGD-derived gene pairs have a higher interaction similarity as compared to SSD-derived gene pairs, especially for older duplicates (Mann–Whitney *U* test; $P < 0.05$). Effect sizes were measured using point-biserial correlation coefficients.

and Methods). We also counted motif frequencies in 1,000 degree-preserving simulated networks, and Z-scores were used to assess the significance of observed motif frequencies. We found that genes derived from (relatively) recent WGD events (up to 30 Ma) are more frequently part of network motifs than genes from more ancient WGD events, demonstrating that motifs are being lost over time (fig. 4*B*). We also found that species with recent WGD events generally have a higher motif frequency than species with more ancient WGDs, regardless of the duplication mode that created the genes forming motifs (e.g., *Z. mays* vs. *O. sativa*; *G. max, A. thaliana, and P. trichocarpa* vs. *V. vinifera* in fig. 4*C*).

Although the *O. sativa* genome has signatures of a WGD event that is shared with all Poaceae, the *Z. mays* genome has undergone an additional WGD event ~19 Ma (table 2 and fig. 4*A*) (Wei et al. 2007; Van de Peer et al. 2017). Likewise, the *V. vinifera* genome has signatures of a WGD event that is shared with all core eudicots, whereas the genomes of *P. trichocarpa, A. thaliana,* and *G. max* have undergone 1, 2, and 2 additional rounds of WGD, respectively (fig. 4*A*) (Van de Peer et al. 2017; Clark and Donoghue 2018). The higher motif frequencies in species with additional WGDs reveal that recent polyploidization events played a key role in the increase of robustness (i.e., greater number of nodes, edges, and interactions thereof) in GRNs, providing genomes with the raw genetic material (i.e., network nodes) for evolving novel regulatory interactions. However, it is noteworthy that such increased motif frequencies were not observed for *S. lycopersicum*, whose genome has signatures of an additional, Solanaceae-specific whole-genome triplication event (fig. 4*B* and *C*). This exceptional case suggests that rates of motif loss vary across clades, and *S. lycopersicum* (or even all Solanaceae) might have faster rates of motif loss.

In human GRNs, Mottes et al. (2021) observed that SSD-derived genes have a greater contribution to the formation of network motifs than WGD-derived genes. However, the two WGD events studied happened around

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

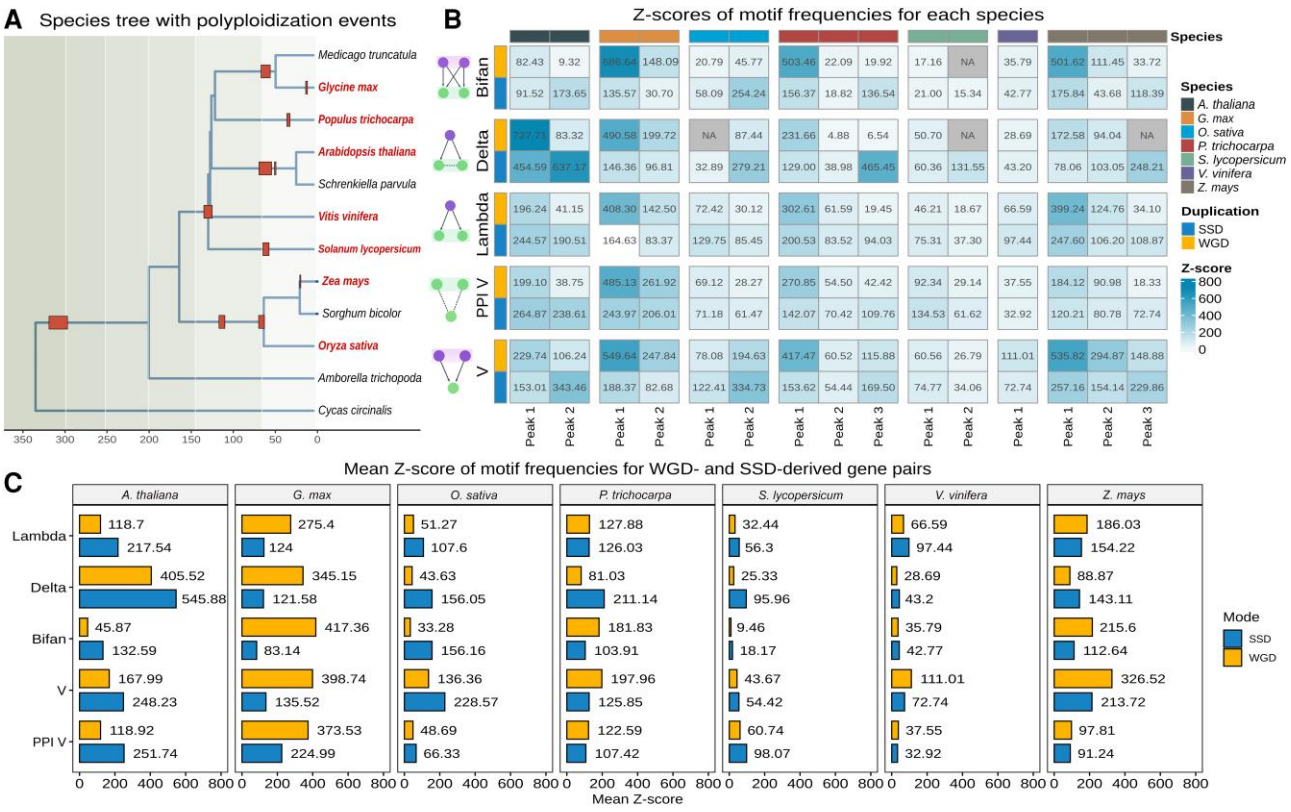

FIG. 4. Z-scores of network motif frequencies in each species. (*A*) Species tree with polyploidization events indicated. Species included in this work are highlighted in bold red text. Red boxes represent polyploidization events. The timetree was obtained from the TimeTree database (Kumar et al. 2022). (*B*) Z-scores of motif frequencies by species and K_s peak-based age group. Z-scores <2 were considered not significant and set to NA. The plot shows that genes derived from recent WGD events are more frequently part of network motifs than genes from more ancient WGD events. (*C*) Mean Z-score of motif frequencies for WGD- and SSD-derived gene pairs. Recent polyploids have higher motif frequencies than ancient polyploids. WGD-derived genes have a greater contribution to motif formation in recent polyploids, whereas SSD-derived genes are more frequent in ancient polyploids.

the emergence of vertebrates (536–525 and 498–486 Ma, respectively) (Simakov et al. 2020; Yu et al. 2023); hence, they are very ancient events. Here, by analyzing WGD events of different ages, we demonstrate that the findings from Mottes et al. (2021) are likely only true for ancient, but not for recent WGD events. We observed that whether WGD or SSD has a greater contribution to motif formation is tightly linked to the WGD background of the species (i.e., ancient vs. recent WGD). WGD-derived gene pairs have a greater contribution to motif formation in species with recent WGD events (e.g., *G. max* and *Z. mays*), whereas SSD-derived genes overtake the contribution of WGD-derived genes in motifs in species with more ancient WGD events (fig. 3*B*). However, it is not clear whether motifs are lost over time because genes involved in motifs are simply lost (because of fractionation) or because networks are rewired.

Our findings demonstrate that WGD and SSD events contributed in different ways to the evolution of GRNs in angiosperms, as recently also observed for human GRNs (Mottes et al. 2021). The greater contribution of recent WGDs to motif formation suggests that WGD events have a more significant impact on the short-term evolution of polyploids. In line with our findings, polyploidy and the short-term redundancy it generates have been associated with an increased robustness against mutations and environment changes (Crow and Wagner 2005; Van de Peer et al. 2009). For instance, WGD events have been correlated with surviving environmental turmoil, such as during the Cretaceous–Paleogene extinction and glaciation events (Vanneste et al. 2014; Van de Peer et al. 2017; Novikova et al. 2018; Yao et al. 2019; Koenen et al. 2021). Assuming motif frequencies as a proxy for GRN robustness, increased chances of survival for recent polyploids could be explained by their increased GRN robustness. Alternatively, as motifs are genetic circuits that perform elementary regulatory functions, the greater number of motifs in recent polyploids creates bigger signals due to the combinatorial effect of each individual motif, as recently demonstrated through simulations of GRN duplications (Ebadi et al. 2023). Ultimately, such increased signals would allow greater jumps in the fitness landscape that would not be possible with SSDs, possibly explaining an increased chance of survival during periods of extinction.

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

Network Motifs Created by WGD and SSD Are Associated With Different Biological Processes

To explore how WGD and SSD shaped angiosperm GRNs from a functional perspective, we performed functional enrichment analyses for GO terms, Interpro domains, and TF families for genes in motifs. Only bifans were enriched in particular functional categories or TF families. Bifans are combinatorial decision-making circuits that integrate different input signals, organizing the transcription of downstream target genes. We found that WGD and SSD contribute to GRNs with genes of different functional groups. WGD-derived genes in bifans are associated with dosage-sensitive processes, whereas SSD-derived genes in bifans are enriched in stress-related processes (fig. 5; supplementary table S2, Supplementary Material online).

WGD-derived bifans are enriched in genes associated with translation, transcriptional regulation, flower development, histone modifications, photosynthesis-related processes (e.g., chlorophyll biosynthesis, thylakoid membrane organization, and RuBisCO), cell cycle regulation, and carbohydrate and lipid metabolism (fig. 5; supplementary table S3, Supplementary Material online). The association of WGD-derived genes with such functional categories has been reported in multiple independent studies (Sato et al. 2009; Conant et al. 2014; Freeling et al. 2015; Almeida-Silva et al. 2020), and it is likely because these gene families are more dosage sensitive than others (Cheng et al. 2018).

The enriched stress-related processes for SSD-derived target genes in bifans include oxidative stress, response to hypoxia, pathogenesis-related proteins (e.g., chitinases and lysozymes, peroxidases, and thaumatins), and recognition of pathogen-associated molecular patterns (e.g., leucine-rich repeat receptor kinase and wall-associated kinases). Likewise, SSD-derived TFs in bifans are enriched in stress-related TF families, such as WRKY, ERF, NAC, and G2-like. Previous studies have reported that stress-related gene families tend to expand through tandem duplications, because their organization in tandem arrays facilitates their coexpression in response to environmental stimuli (Amoutzias and Van de Peer 2008; Hanada et al. 2008; Rodgers-Melnick et al. 2012). Although WGD is the main force driving the expansion of TFs due to dosage balance, hence increasing the number of nodes in GRNs, our findings demonstrate that SSD also plays an important role in GRNs by contributing with stress-related genes.

Collectively, the patterns we observe for WGD- and SSD-derived motifs in this study are very similar to what has been observed for WGD- and SSD-derived individual genes. In summary, we show that: 1) WGD-derived genes and motifs are lost over time, but some are retained over long evolutionary timescales; 2) WGD-derived genes and motifs that are retained long-term are associated with growth and development, with a particular preferential retention of dosage-sensitive molecular pathways, such as transcriptional regulation, histone modifications, and cell cycle regulation; and 3) SSD-derived genes and motifs that are retained long(er) term are associated with stress-related functions. Such similarities suggest that the evolution of motifs is intricately associated with the evolution of genes, and a more detailed picture could be observed by comparing rates of gene and motif loss.

For instance, we observed that most WGD-derived motifs are quickly lost over time (within a few tens of My) as compared with SSD-derived motifs (see previous section, fig. 4*B*). On the one hand, the rapid loss of WGD-derived motifs could be explained by the rapid loss of WGD-derived genes due to fractionation (Freeling et al. 2015). On the other hand, if rates of motif loss were higher than rates of gene loss, it would suggest that motifs are lost first (i.e., due to transcriptional rewiring), leading to relaxed selection pressures that could eventually lead to gene loss or, in special cases, to subfunctionalization and/or neofunctionalization. The latter hypothesis implies that participating in motifs imposes additional selective constraints on genes, and hence, motifs could have a significant long-term impact on evolution.

Additionally, as motifs are subgraphs that are observed more often than is expected by chance, they are thought to have been under positive selection (Mottes et al. 2021). We hypothesize that selection acting upon motifs leads to a higher chance of retention for genes that participate thereof. In the long run, transcriptional rewiring of motifs created by WGD events could contribute to evolutionary innovations and increase in organismal complexity, such as what has been observed for the auxin response regulators AUX/IAA and some subfamilies of MADS-box TFs in plants (Remington et al. 2004; Veron et al. 2007), and the RAR/RXR pathway and homeobox (Hox) gene clusters in vertebrates (Holland et al. 2008; Mottes et al. 2021).

Conclusions

In this paper, we explored the impact of WGD and SSD events on the topology of GRNs in angiosperms. We observed that each duplication mechanism contributed to network evolution in its own way. Using a diverse set of species with different numbers and ages of WGD, we confirm previous observations that WGD-derived genes are under strong selective constraints, have slower rates of sequence divergence, and tend to preserve interactions with themselves and with other partners. These findings provide robust evidence that previously reported patterns for WGD-derived genes are universal, regardless of the number and age of genome duplications. WGD events contribute to GRNs with genes associated with dosage-sensitive processes, whereas SSD contributes with stress-related genes. Strikingly, recent polyploids have higher motif frequencies than ancient polyploids, suggesting that polyploidy might have a great(er) impact on the short term. We hypothesize that WGD events confer a short-term selective advantage to polyploids by enhancing

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

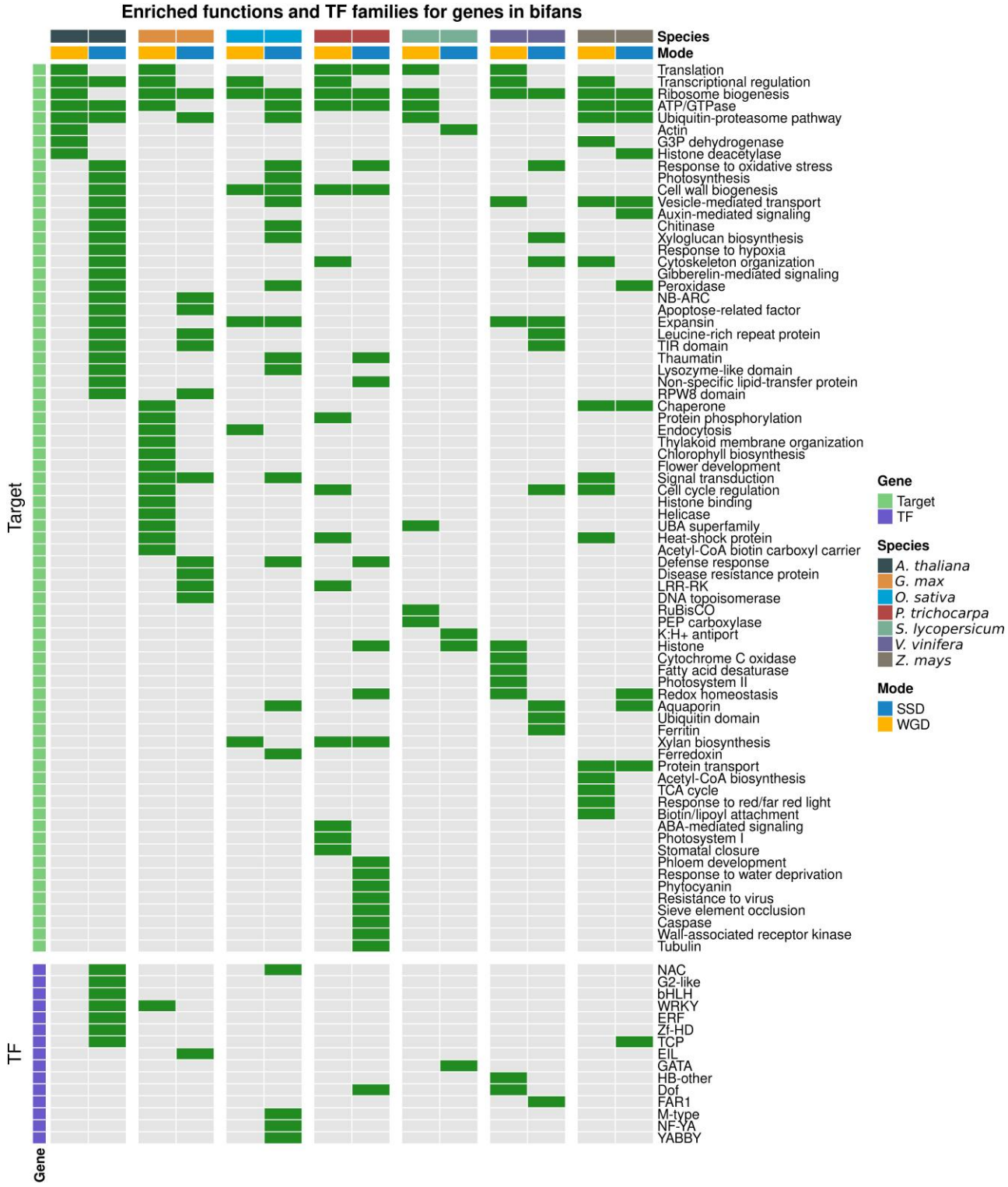

Fig. 5. Enrichment of transcription factor families and gene functions for WGD- and SSD-derived genes in bifans. Enrichment analyses for TF families in V and bifans were performed using all TFs in the GRN as background. Green and gray cells represent enrichment and no enrichment, respectively. Enrichment analyses for functional categories in lambda, delta, and bifans were performed using all targets in the GRN as background. WGD-derived genes in bifans are enriched in dosage-sensitive processes, such as transcriptional regulation, regulation of cell cycle, translation, and photosynthesis, whereas SSD-derived genes in bifans are enriched in stress-related processes, such as response to hypoxia and oxidative stress, pathogenesis-related proteins, recognition of pathogen-associated molecular patterns, and stress-related TF families, such as ERF, NAC, and WRKY.

GRN robustness and potentially magnifying GRN output signals (e.g., traits, gene expression, etc.) through the

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

combinatorial effect of each individual motif, as it increases the frequencies of all motif types, especially complex motifs such as bifans. Such magnified signals can result in greater jumps in the fitness landscape and increased chances of survival during periods of environmental instability (see also Ebadi et al. 2023). Our findings underscore the importance of WGD events in the evolution of GRNs and highlight their potential role in facilitating adaptation to changing environments as a result of greater GRN robustness.

Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Y.V.d.P. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522). Y.V.d.P. and F.A.-S. acknowledge funding from Ghent University (Methusalem funding, BOF.MET.2021.0005.01).

Data Availability

All data and code used in this paper are available in a GitHub repository (https://github.com/almeidasilvaf/polyploid_GRNs) to ensure full reproducibility.

References

Almeida-Silva F, Moharana KC, Machado FB, Venancio TM. 2020. Exploring the complexity of soybean (glycine max) transcriptional regulation using global gene co-expression networks. *Planta* **252**:1–12.

Almeida-Silva F, Venancio TM. 2022. BioNERO: an all-in-one R/bioconductor package for comprehensive and easy biological network reconstruction. *Funct Integr Genomics*. **22**:131–136.

Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**:1711–1714.

Amoutzias G, Van de Peer Y. 2008. Together we stand: genes cluster to coordinate regulation. *Dev Cell*. **14**:640–642.

Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol*. **186**:54–62.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell*. **16**:1679–1691.

Burda Z, Krzywicki A, Martin OC, Zagorski M. 2011. Motifs emerge from function in model gene regulatory networks. *Proc Natl Acad Sci U S A*. **108**:17263–17268.

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants*. **4**:258–268.

Clark JW, Donoghue PC. 2017. Constraining the timing of whole genome duplication in plant evolutionary history. *Proc R Soc B Biol Sci*. **284**:20170912.

Clark JW, Donoghue PC. 2018. Whole-genome duplication and plant macroevolution. *Trends Plant Sci*. **23**:933–945.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet*. **6**:836–846.

Conant GC, Birchler JA, Pires JC. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol*. **19**:91–98.

Crow KD, Wagner GP. 2005. What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol*. **23**:887–892.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal Complex Syst*. **1695**:1–9.

Defoort J, Van de Peer Y, Carretero-Paulet L. 2019. The evolution of gene duplicates in angiosperms and the impact of protein–protein interactions and the mechanism of duplication. *Genome Biol Evol*. **11**:2292–2305.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*. **3**:e314.

De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci U S A*. **110**:2898–2903.

De Smet R, Van de Peer Y. 2012. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr Opin Plant Biol*. **15**:168–176.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nat Protoc*. **4**:1184–1191.

Ebadi M, Bafort Q, Mizrachi E, Audenaert P, Simoens P, Van Montagu M, Bonte D, Van de Peer Y. 2023. The duplication of genomes and gene regulatory networks and its potential for evolutionary adaptation and survival. *bioRxiv*:2023–2004.

Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW. 2013. The roles of whole-genome and small-scale duplications in the functional specialization of Saccharomyces cerevisiae genes. *PLoS Genet*. **9**:e1003176.

Finn RD, Clements J, Eddy SR. 2011. HMMER Web server: interactive sequence similarity searching. *Nucleic Acids Res*. **39**:W29–W37.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol*. **60**:433–453.

Freeling M, Scanlon MJ, Fowler JE. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr Opin Genet Dev*. **35**:110–118.

Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res*. **16**:805–814.

Gera T, Jonas F, More R, Barkai N. 2022. Evolution of binding preferences among whole-genome duplicated transcription factors. *Elife*. **11**:e73225.

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol*. **148**:993–1003.

Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez È, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res*. **18**:1100–1111.

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring regulatory networks from expression data using tree-based methods. *PloS One*. **5**:e12776.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**:772–780.

Koenen EJ, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Herendeen PS, Bruneau A, Hughes CE. 2021. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the cretaceous–paleogene (K–pg) mass extinction event. *Syst Biol*. **70**:508–526.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. Timetree 5: an expanded resource for species divergence times. *Mol Biol Evol*. **39**:msac174.

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am J Bot*. **105**:348–363.
Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. 2016. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell*. **28**:326–344.
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*. **102**:5454–5459.
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002. Network motifs: simple building blocks of complex networks. *Science* **298**:824–827.
Moharana KC, Venancio TM. 2020. Polyploidization events shaped the transcription factor repertoires in legumes (Fabaceae). *Plant J*. **103**:726–741.
Mottes F, Villa C, Osella M, Caselle M. 2021. The impact of whole genome duplications on the human gene regulatory networks. *PLoS Comput Biol*. **17**:e1009638.
Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B. 2021. Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat Commun*. **12**:1–14.
Novikova PY, Hohmann N, Van de Peer Y. 2018. Polyploid Arabidopsis species originated around recent glaciation maxima. *Curr Opin Plant Biol*. **42**:8–15.
Ohno S. 1970. *Evolution by gene duplication*. Heidelberg: Springer Science & Business Media.
Panchy NL, Azodi CB, Winship EF, O'Malley RC, Shiu S-H. 2019. Expression and regulatory asymmetry of retained *Arabidopsis thaliana* transcription factor genes derived from whole genome duplication. *BMC Evol Biol*. **19**:1–17.
Papatheodorou I, Moreno P, Manning J, Fuentes AM-P, George N, Fexova S, Fonseca NA, Füllgrabe A, Green M, Huang N. 2020. Expression atlas update: from tissues to single cells. *Nucleic Acids Res*. **48**:D77–D83.
Patil I. 2021. Visualizations with statistical details: the "ggstatsplot" approach. *J Open Source Softw*. **6**:3167.
Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol*. **20**:1–23.
Remington DL, Vision TJ, Guilfoyle TJ, Reed JW. 2004. Contrasting modes of diversification in the aux/IAA and ARF gene families. *Plant Physiol*. **135**:1738–1752.
Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant*. **11**:414–428.
Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, DiFazio SP. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in Populus. *Genome Res*. **22**:95–105.
Sato Y, Hashiguchi Y, Nishida M. 2009. Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. *BMC Evol Biol*. **9**:1–14.
Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr Opin Plant Biol*. **15**:147–153.
Simakov O, Marlétaz F, Yue J-X, O'Connell B, Jenkins J, Brandt A, Calef R, Tung C-H, Huang T-K, Schmutz J. 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol*. **4**:820–830.
Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. **49**:D605–D612.
Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol*. **207**:454–467.
Tasdighian S, Van Bel M, Li Z, Van de Peer Y, Carretero-Paulet L, Maere S. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell*. **29**:2766–2785.
Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res*. **13**:382–390.
Teufel AI, Johnson MM, Laurent JM, Kachroo AH, Marcotte EM, Wilke CO. 2019. The many nuanced evolutionary consequences of duplicated genes. *Mol Biol Evol*. **36**:304–314.
Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G. 2020. Plantregmap: charting functional regulatory maps in plants. *Nucleic Acids Res*. **48**:D1104–D1113.
Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. **33**:11–26.
Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet*. **10**:725–732.
Van de Peer Y, Maere S, Meyer A. 2010. 2R Or not 2R is not the question anymore. *Nat Rev Genet*. **11**:166–166.
Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nat Rev Genet*. **18**:411–424.
Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the cretaceous–paleogene boundary. *Genome Res*. **24**:1334–1347.
Veron AS, Kaufmann K, Bornberg-Bauer E. 2007. Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. *Mol Biol Evol*. **24**:670–678.
Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. Kaks_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinform*. **8**:77–80.
Ward JJ, Thornton JM. 2007. Evolutionary models for formation of network motifs and modularity in the Saccharomyces transcription factor network. *PLoS Comput Biol*. **3**:e198.
Wei F, Coe ED, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S. 2007. Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet*. **3**:e123.
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**:708–713.
Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol*. **8**:e1000409.
Yao Y, Carretero-Paulet L, Van de Peer Y. 2019. Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy. *PLoS One*. **14**:e0220257.
Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI, Carbajo Martinez M, Chakiachvili M, *et al.* 2022. Ensembl genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res*. **50**:D996–D1003.
Yu D, Ren Y, Uesaka M, Beavan AJ, Muffato M, Shen J, Li Y, Sato I, Wan W, Clark JW. 2023. Hagfish genome illuminates vertebrate whole genome duplications and their evolutionary consequences. *bioRxiv* 2023–2004.

Downloaded from https://academic.oup.com/mbe/article/40/7/msad141/7219705 by guest on 29 August 2023